INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Approximations in dynamic zero-sum games, II

Mabel. TIDBALL, Odile. POURTALLIER and Eitan. ALTMAN

N° 2348

Septembre 1994

PROGRAMME 5

Traitement du signal,

automatique

et productique

ISSN 0249-6399

Approximations in dynamic zero-sum games, II

Mabel. TIDBALL, Odile. POURTALLIER and Eitan. ALTMAN *

Programme 5 — Traitement du signal, automatique et productique
Projet MIAOU

Rapport de recherche n°2348 — Septembre 1994 — 24 pages

Abstract: We pursue in this paper our study of approximations of values and ϵ-saddle-point policies in dynamic zero-sum games. After extending the general theorem for approximation, we study zero-sum stochastic games with countable state space, and non-bounded immediate reward. We focus on the expected average payoff criterion. We use some tools developed in the first paper, to obtain the convergence of the values as well as the convergence of the ϵ saddle-point policies in various approximation problems. We consider several schemes of truncation of the state space (e.g. finite state approximation) and approximations of games with discount factor close to one by the game with expected average cost. We use the extension of the general Theorem for approximation to study approximations in stochastic games with complete information. We finally consider the problem of approximating the sets of policies. We obtain some general results that we apply to a pursuit evasion differential game.

(Résumé : tsvp)

*INRIA, Centre Sophia-Antipolis, 2004 Route des Lucioles, B.P.93, 06902 Sophia-Antipolis Cedex France

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Approximations dans les jeux dynamiques à somme nulle, II

Résumé : Nous poursuivons dans ce papier une étude portant sur l'approximation de la fonctions valeur, ainsi que des strategies ϵ-optimal pour des jeux dynamiques à deux joueurs et à somme nulle.

Nous étendons dans un premier temps un théorème général utilisé pour les approximations, puis nous étudions des jeux stochastiques à somme nulle dont l'espace d'état est dénombrable, et dont le coût instantané est non borné. On s'interesse plus particulièrement au coût moyen. Nous utilisons des outils que nous avons developpés dans le précédent papier, pour obtenir la convergence de la fonction valeur ainsi que la convergence des stratégies ϵ-optimales dans différents problèmes d'approximation.

Nous considérons différents types d'approximations d'espaces d'états infinis par des espaces d'état finis ainsi que des approximations de jeux avec des taux d'actualisation proche de 1, par des jeux à coûts moyens.

Nous utilisons l'extension du théorème général pour les approximations pour étudier des approximations pour des jeux stochastiques avec information complète.

On considère enfin le problème de l'approximation fini des ensembles de stratégies. Nous obtenons des résultats généraux que nous appliquons au cas de jeux différentiels de poursuite-évasion.

1 Introduction

We pursue in this paper our study of approximations of values and saddle-point policies in dynamic zero-sum games. In a previous paper [25], we developed some tools for approximating zero-sum games, and applied them to stochastic games with discounted payoff criterion. In this paper we extend the general theory for approximation to handle cases where a value does not exist for the limit game, and we apply the general theorems for approximation to the following dynamic zero-sum games.

We first consider approximation problems arrizing in stochastic games with expected average cost: finite state approximation of stochastic games with a countable state space, and convergence of stochastic games with discounted cost to the stochastic game with average cost. We then consider approximations in stochastic games with complete information, and problems in dynamic games related to discretizing of the strategy sets.

There is a rich literature on finite state approximation in the context of a single controller. The discounted reward was extensively studied, see [2, 11, 16, 17, 26, 27], and [20, 28, 29] for related discretization results. For the expected average cost, there exist only few work on state approximations in the context of control, and none in the context of stochastic games. Even if existing schemes could be extended to the setting of stochastic game, they are still quite restricted since their convergence (in the setting of control) was established under conditions that seem very strong, and quite often non-applicable. Thomas and Stengos obtained several schemes for finite state approximations. They impose, some scrambling conditions which should hold uniformly in the states. They do not seem to hold for queueing applications, such as the models in [3, 4, 6]. Altman introduced several finite state approximation schemes [1, 2] for constrained control. They do not require the scrambling conditions, but have other restrictive conditions: the scheme in [1] requires some monotone structure on the immediate cost, and holds for immediate costs that are only functions of the state, and not of the actions. The scheme in [2] has the "finite neighbor" restriction, i.e., from each state, only finitely many states are accessible within one step.

The first approximation scheme that we introduce in the current paper relaxes the above restrictions, and is thus also useful and new in the case of a single controller.

The second scheme which we propose in this paper is an adaptation of the scheme from [2]. In both schemes, in addition to the convergence of the value, which is the question studied in most of the papers on state approximations, we obtain (i) the convergence of the policies, (ii) the robustness of policies, i.e. an equilibrium point for the limiting (infinite state) stochastic game $G = G_\infty$ is shown to be ϵ-equilibrium

for the approximating games G_n with all n large enough. On the other hand, for any ϵ, the equilibrium policies for G_n are almost optimal for the limiting game, for all n large enough.

In the previous paper we focused on approximations of stochastic games with discounted cost and bounded reward, and mentioned that standard techniques can be used to transform problems with unbounded reward to problems with bounded ones. This is, however, not the case for the expected average payoff criterion. The question of existence of value and of equilibrium stationary policies (under some recurrence conditions) for the case of unbounded reward was solved recently in [6, 7, 10, 21]. The growing interest in stochastic games with unbounded cost in recent years was partly driven by applications of stochastic games in telecommunications systems in general, and in queueing systems in particular. Although queues are always finite in practice (which results in a finite state space description), models of infinite queues are frequently more useful since they are usually easier to solve. Indeed, several dynamic games arising in such applications were explicitly solved [6, 8], or, at least reduced to the search for equilibrium policies among small classes of policies [4, 5, 6]. The scheduling problem described in [6, 8], the routing problem into two queues [4, 6], the flow and service control in [5] have not been solved for the case of finite state space, since there is an effect of the boundaries due to the finiteness of the queues that destroys the nice structure of the problem with infinite state space. In all the above problems, it is unnatural to consider bounded costs. Since costs represent queue lengths or waitinig times, these typically grow to infinity as the number of "customers" in the queues grows to infinity. The theory developed in this paper allows to use the equilibrium policies obtained for the infinite queues to construct ϵ-equilibrium policies for the corresponding problems with finite queues, provided they are sufficiently large.

A second issue in this paper is the convergence of stochastic games in the discount factor. The convergence of the value and equilibrium policies for discounted cost stochastic games to those of the average cost game are well known, see e.g. [14]. These were extended recently to unbounded cost (see [7, 21]). We not only obtain an alternative proof for the above convergence of the values and policies, but also obtain new robustness results.

When the players are restricted to use pure strategies in a stochastic game, the game in general does not have a value anymore. Using an extension of the general approximation theorems, we study approximations under that restriction. This yields approximation theorems for stochastic games with complete information (where player 2 knows at time t the action taken by player 1 at time t).

Finally, we consider the problem of approximating the set of policies by other sets. We obtain a general approximating theorem for the case that the strategy sets are endowed with the Hausdorf metric. We apply the theorem to a zero-sum pursuit evasion differential game introduced in [9, 22].

The structure of the paper is the following. We begin by citing and extending the general theory for approximations, developed in [25], in Section 2. We then introduce in Section 3 the model, notation and assumptions for the stochastic game. We present two schemes for state approximation in Section 4. The convergence in the discount factor is established in Section 5. In Section 6 we discuss approximations for stochastic games with complete information. The approximation of the strategy sets is finially presented in Section 7 together with the application to the pursuit evasion game.

2 Key Theorems for approximations

We consider the following sequence $G_n = (S_n, U_n, V_n)$ $n = 1, 2, ..., \infty$ of generic zero-sum games where U_n is the set of strategies (or policies) of player one and V_n is the set of strategies of player two for the nth game. We assume that both U_n and V_n are endowed with some topology. $S_n : U_n \times V_n \to \mathbb{R}$ is a measurable function for all n. We define the upper (lower) value of the game:

$$\overline{R_n} = \inf_{v \in V_n} \sup_{u \in U_n} S_n(u, v) \quad \left(\underline{R_n} = \sup_{u \in U_n} \inf_{v \in V_n} S_n(u, v) \right). \tag{1}$$

$G = (S, U, V) \stackrel{\text{def}}{=} (S_\infty, U_\infty, V_\infty)$ will be called the limit game. It will be assumed that it has a value $R \stackrel{\text{def}}{=} R_\infty = \mathbf{Val}\{S(u, v)\}_{u,v}$.

A strategy $u^* \in U_n$ is said to be ϵ-optimal for player one in game n if

$$\inf_{v \in V_n} S_n(u^*, v) \geq \inf_{v \in V_n} S_n(u, v) - \epsilon \qquad \forall u \in U_n \tag{2}$$

which is equivalent to $\inf_{v \in V_n} S_n(u^*, v) \geq \underline{R}_n - \epsilon$. It is said to be strong ϵ-optimal for player one in game n if it satisfies

$$\inf_{v \in V_n} S_n(u^*, v) \geq \overline{R}_n - \epsilon$$

A strategy $v^* \in V_n$ is said to be ϵ-optimal for player two in game n if

$$\sup_{u \in U_n} S_n(u, v^*) \leq \sup_{u \in U_n} S_n(u, v) + \epsilon \qquad \forall v \in V_n \tag{3}$$

which is equivalent to $\sup_{u\in U_n} S_n(u,v^*) \leq \overline{R}_n + \epsilon$. It is said to be strong ϵ-optimal if

$$\sup_{u\in U_n} S_n(u,v^*) \leq \underline{R}_n + \epsilon$$

Note that strongly ϵ-optimality implies ϵ-optimality. If a game has a value $\underline{R}_n = \overline{R}_n$ then strong ϵ-optimality is equivalent to ϵ-optimality.

Assume that (S_n, U_n, V_n) converge (in some sense) to (S, U, V). We are interested in the following questions:
(Q1) Convergence of the values: does $\underline{R}_n$ (or $\overline{R}_n$) converge to R?
(Q2) Convergence of policies: Fix some $\epsilon \geq 0$. Let ϵ_n be a sequence of positive real numbers such that $\overline{\lim}_{n\to\infty} \epsilon_n \leq \epsilon$. Assume that u_n^* and v_n^* are ϵ_n-optimal policies for the nth game. Are u_n^* and v_n^* "almost" optimal for the limit game, for all n large enough?
(Q3) Let $\overline{u} \in U$ (resp. $\overline{v} \in V$) be some limit point of u_n^* (resp. v_n^*), defined above. Is $\overline{u}$ (resp. $\overline{v}$) ϵ-optimal for the limit game?
(Q4) Robustness of the optimal policy: If u^* (resp. v^*) is an ϵ-optimal for the limit game, can we derive of it an "almost" (strong) optimal policy for the nth approximating game, for all n large enough?

A straightforward generalization of Theorem 2.1 in [25] yields:

Theorem 2.1 *Assume that for any $\epsilon_1 > 0$ there exists a sequence of functions, $\pi_n^1 : U_n \to U$, $\pi_n^2 : V_n \to V$, $\sigma_n^1 : U \to U_n$, $\sigma_n^2 : V \to V_n$, $n = 1, 2, ...$ such that*
($\mathcal{A}1$): $\overline{\lim}_{n\to\infty}[S_n(u, \sigma_n^2(v)) - S(\pi_n^1(u), v)] \leq \epsilon_1$ uniformly in $u \in U_n$ for each $v \in V$.
($\mathcal{A}2$): $\underline{\lim}_{n\to\infty}[S_n(\sigma_n^1(u), v) - S(u, \pi_n^2(v))] \geq -\epsilon_1$ uniformly in $v \in V_n$ for each $u \in U$.
Then
(1) $\lim_{n\to\infty} \underline{R}_n = \lim_{n\to\infty} \overline{R}_n = R$.
(2) For any $\epsilon' > \epsilon + 3\epsilon_1$, there exists N such that $\pi_n^1(u_n^*)$ (resp. $\pi_n^2(v_n^*)$, see definitions in (Q2)) is ϵ'-optimal for the limit game, for all $n \geq N$.
(3) Let u^* (resp. v^*) be ϵ-optimal for the limit game. Then for all $\epsilon' > \epsilon + 3\epsilon_1$, there exists $N(\epsilon')$ such that $\sigma_n^1(u^*)$ (resp. $\sigma_n^2(v^*)$) is strong ϵ'-optimal for the nth approximating game, for all $n \geq N(\epsilon')$.
(4) Suppose
($\mathcal{A}3$): $S(u, v)$ is a lower semicontinuous function in u,
($\mathcal{A}4$): $S(u, v)$ is an upper semicontinuous function in v.
Suppose $\bar{u} \in U$ (resp. $\bar{v} \in V$) is a limit point of $\pi_n^1(u_n^*)$ (resp. $\pi_n^2(v_n^*)$). Then $\bar{u}$ (resp. $\bar{v}$) is $(\epsilon + 5\epsilon_1)$-optimal for the limit game.

Remark 2.1 *(i) Whenever $U_n = U$ and $V_n = V$ do not depend on n, π_n and σ_n will be chosen as the indentity maps.*
(ii) It follows from the proof of part (1) in the above Theorem that if for all G_n, $n = 1, 2, ..., \infty$ there exist optimal policies for both players and if $U_n = U$ and $V_n = V$ do not depend on n, then

$$|\overline{R}_n - R| \leq \sup_{u,v} |S_n(u,v) - S(u,v)|, \qquad |\underline{R}_n - R| \leq \sup_{u,v} |S_n(u,v) - S(u,v)|$$

Next, we relax the assumption that the limit game has a value: $\underline{R}_\infty \neq \overline{R}_\infty$. We show that Theorem 2.1 still holds, by appropriately enlarging the policy spaces and redefining the cost, so that the upper (or lower) value becomes a real value of a new game.

We consider the convergence of the upper values (and corresponding optimal or almost optimal policies) of the approximating games to those of the limit game. The corresponding convergence for the lower values are obtained in the same way. Define $\mathcal{U}_n = \{$ the class of functions $U_n \to V_n\}$. Define the cost $\hat{\mathcal{S}}_n : \mathcal{U}_n \times V_n \to \mathbb{R}$ by $\hat{\mathcal{S}}_n(\psi, v) = S_n(\psi(v), v)$.

Lemma 2.1 *(i) For all n, the new game $\mathcal{G}_n = (\hat{\mathcal{S}}_n, \mathcal{U}_n, V_n)$ has a value $\mathcal{R}_n$, and $\mathcal{R}_n = \overline{R}_n$.*
(ii) v^* is $\epsilon-$optimal for player 2 in game $\mathcal{G}_n$ if and only if it is $\epsilon-$optimal in game G_n.

Proof.

$$\sup_{v \in V_n} \inf_{\psi \in \mathcal{U}_n} \hat{\mathcal{S}}(\psi, v) = \sup_{v \in V_n} \inf_{\psi \in \mathcal{U}_n} S(\psi(v), v) = \sup_{v \in V_n} \inf_{u \in U_n} S(u, v).$$

On the other hand,

$$\sup_{v \in V_n} \inf_{\psi \in \mathcal{U}_n} S(\psi(v), v) = \inf_{\psi \in \mathcal{U}_n} \sup_{v \in V_n} S(\psi(v), v) = \inf_{\psi \in \mathcal{U}_n} \sup_{v \in V_n} \hat{\mathcal{S}}(\psi, v).$$

(i) is obtained by combining the last two equations. (ii) follows since for any $v \in V$,

$$\sup_{u \in U_n} S_n(u, v) = \sup_{\psi \in \mathcal{U}_n} \hat{\mathcal{S}}_n(\psi, v).$$

■

By using the new games for which the values exist, and applying Theorem 2.1, we may conclude the following convergence properties of the original games.

Theorem 2.2 *Assume that the functions π_n and σ_n exist as in Theorem 2.1, and that conditions ($\mathcal{A}1$) and ($\mathcal{A}2$) hold. Then*
(1) $\lim_{n\to\infty} \overline{R}_n = \overline{R}$, $\lim_{n\to\infty} \underline{R}_n = \underline{R}$.
(2) For any $\epsilon' > \epsilon + 3\epsilon_1$, there exists N such that $\pi_n^1(u_n^*)$ (resp. $\pi_n^2(v_n^*)$, see definitions in (Q2)) is ϵ'-optimal for the limit game, for all $n \geq N$.
(3) Let u^* (resp. v^*) be ϵ-optimal for the limit game. Then for all $\epsilon' > \epsilon + 3\epsilon_1$, there exists $N(\epsilon')$ such that $\sigma_n^1(u^*)$ (resp. $\sigma_n^2(v^*)$) is ϵ'- optimal for the n approximating game, for all $n \geq N(\epsilon')$.

Proof. Consider the new games $\mathcal{G}_n$ defined above. We show that the assumptions of Theorem 2.1 hold also for $\mathcal{G}_n$. The mapping $\tilde{\pi}_n^2, \tilde{\sigma}_n^2$ for the new games are unchanged;

$$\tilde{\pi}_n^2 = \pi_n^2, \qquad \tilde{\sigma}_n^2 = \sigma_n^2.$$

The mappings $\tilde{\pi}_n^1 : \mathcal{U}_n \to \mathcal{U}$ and $\tilde{\sigma}_n^1 : \mathcal{U} \to \mathcal{U}_n$ for the new games are defined as

$$[\tilde{\pi}_n^1(\psi)](v) = \pi_n^1(\psi(v)), \qquad \forall v \in V, [\tilde{\sigma}_n^1(\psi)](v) = \sigma_n^1(\psi(v)), \qquad \forall v \in V_n.$$

With these definitions as well as the definition of the costs $\hat{\mathcal{S}}_n$, it follows that ($\mathcal{A}1$) and ($\mathcal{A}2$) hold for $\mathcal{G}_n$. The proof now follows by Lemma 2.1. ∎

We may further obtain convergence results for the optimal (or ϵ-optimal) *responses* (in case that the value of the limit game does not exist). To simplify the formulation, this is done below in terms of the new games $\mathcal{G}_n$:

Theorem 2.3 *Consider the new games $\mathcal{G}_n$, and let ψ_n^*, v_n^* be defined as u_n^* in (Q2) (above Theorem 2.1. Under the conditions of Theorem 2.2,*
(1) $\lim_{n\to\infty} \mathcal{R}_n = \mathcal{R} = \overline{R}$,
(2) For any $\epsilon' > \epsilon + 3\epsilon_1$, there exists N such that $\tilde{\pi}_n^1(\psi_n^*)$ (resp. $\tilde{\pi}_n^2(v_n^*)$) is $\epsilon'-$optimal for $\mathcal{G}_\infty$, for all $n \geq N$.
(3) Let ψ^* be ϵ-optimal for player 1 in the limit game $\mathcal{G}_\infty$. Then for all $\epsilon' > \epsilon + 3\epsilon_1$, there exists $N(\epsilon')$ such that $\tilde{\sigma}_n^1(\psi^*)$ (resp. $\tilde{\sigma}_n^2(v^*)$) is ϵ'- optimal for the n approximating game $\mathcal{G}_n$ for all $n \geq N(\epsilon')$.

Next, we consider the result corresponding to statement (4) in Theorem 2.1.

Theorem 2.4 *Assume that the conditions of Theorem 2.3 hold, that the set of response-strategies for player 2 in games $\mathcal{G}_n$ is endowed with some topology, and that ($\mathcal{A}3$) and ($\mathcal{A}4$) hold for game $\mathcal{G}_\infty$. Then statement (4) of Theorem 2.1 holds for games $\mathcal{G}_n$.*

3 Stochastic games with expected average payoff

We consider the two person, zero sum stochastic game defined by the objects $\{\mathbf{I}, \mathbf{A}, \mathbf{B}, P, r\}$, where

- $\mathbf{I}$ is a countable state space;
- $\mathbf{A}$ and $\mathbf{B}$ are sets of actions for player I and player II, respectively; at each state $j \in \mathbf{I}$, the available actions for the players are $\mathbf{A}_j$ and $\mathbf{B}_j$ respectively. These sets are assumed to be compact metric sets.
- $P(a,b) = [p(i,a,b,j)]_{i,j}$, $a \in \mathbf{A}$, $b \in \mathbf{B}$, $i,j \in \mathbf{I}$ are the transition probabilities, so that $p(i,a,b,j)$ is the probability to move from i to j if the players use actions a and b.
- $r : \mathbf{I} \times \mathbf{A} \times \mathbf{B} \to \mathbb{R}$ is an immediate reward function.

The game is played in stages $t = 0, 1, 2, ...$ If at some stage t the state is i, then the players independently choose actions $a \in \mathbf{A}_i$, $b \in \mathbf{B}_i$. Player II then pays player I the amount $r(i,a,b)$ and at stage $t+1$ the new state is chosen according to the transition probabilities $p(i,a,b,\bullet)$. The game continues at this new state.
Let U and V be the set of behavioral strategies for both players. A strategy $u \in U$ is a sequence $u = (u_0, u_1, ...)$ where u_t is a probability measure over the available actions, given the whole history of previous states and of previous actions of both players as well as the current state.

A Markov policy $q = \{q_0, q_1, ...\}$ is a policy (for either player one or two) where q_t is allowed to depend only on t and on the state at time t.

A *stationary (mixed) policy* g for player one is characterized by a conditional distribution
$p^g(\bullet \mid j)$ over $\mathbf{A}_j$, so that $p^g(\mathbf{A}_j \mid j) = 1$, which is interpreted as the distribution over the actions available at state j which player I uses when it is in state j. With some abuse of notation, we shall set $g(\bullet \mid j) = p^g(\bullet \mid j)$ for stationary g. Let S^A be the set of stationary policies for player 1, and define similarly the stationary policies S^B for player 2. If both players use stationary policies, say u and v, then $\{X_t\}$ becomes a Markov chain with stationary transition probabilities, given by

$$p(j,u,v,k) = \int_{\mathbf{A}_j} \int_{\mathbf{B}_j} p(j,a,b,k) u(da|j) v(db|j). \tag{4}$$

Denote $P(u,v) = [p(j,u,v,k)]_{j,k}$.

Next, we introduce a metric topology on the sets of stationary policies. For any compact metric set Γ, let $M_1(\Gamma)$ denote the set of probability measures on the Borel subsets of Γ endowed with the weak topology $\xi(\Gamma)$ (see [19]). The class of stationary policies for player 1 (and similarly for player 2) can be identified with the set $\prod_{i\in\mathbf{I}} M_1(\mathbf{A}_i) \times M_1(\mathbf{B}_i)$; moreover it is compact with respect to the product topology $\prod_{i\in\mathbf{I}} \xi(\mathbf{A}_i) \times \xi(\mathbf{B}_i)$.

Let (u, v) be a pair of strategies and let $i \in \mathbf{I}$ be a fixed initial state. Let $I_t, A_t, B_t, t = 0, ...$ be the resulting stochastic process of the states and actions of the players. Let $E_i^{u,v}$ denote the expectation with respect to the measure defined by u, v, i.

Let $\mu : \mathbf{I} \to \mathbb{R}$ be some positive function. Following Dekker and Hordijk [13] and Spieksma [23], define the μ-norm of any vector $x \in \mathbb{R}^{\mathbf{I}}$ as

$$\| x \|_\mu = \sup_{i\in\mathbf{I}} \frac{|x_i|}{\mu_i}.$$

In a similar way we will use the μ-J-norm, for any finite subset J of the state space E, defined by

$$\| x \|_\mu^J = \sup_{i\in J} \frac{|x_i|}{\mu_i}.$$

Define the μ-norm of matrices $Q \in \mathbb{R}^{\mathbf{I}\times\mathbf{I}}$ as

$$\|Q\|_\mu = \sup_{i\in\mathbf{I}} \mu_i^{-1} \sum_{j\in\mathbf{I}} |Q_{ij}|\mu_j$$

We denote $\bigvee_\mu$ the space of all vectors that are $\mu-$ bounded.

Introduce the following assumptions:

- $(\mathcal{B}1)$
  - i) The instantaneous reward $r(i, a, b)$ is continuous and μ - bounded, i.e.;

    $$\sup_{i\in\mathbf{I}} \sup_{a,b} \frac{|r(i, a, b)|}{\mu_i} \leq M < +\infty$$

    This condition can be rewritten as $\|r(\cdot, u, v)\|_\mu \leq M < +\infty$ for all pure stationary policies u and v.
  - ii) The transition probabilities are μ continuous, i.e. if $a(n) \to a$, $b(n) \to b$ when $n \to +\infty$ then:

    $$\lim_{n\to\infty} \sum_{j\in\mathbf{I}} |p(i, a(n), b(n), j) - p(i, a, b, j)|\, \mu_j = 0$$

- $(\mathcal{B}2)$
  - i) Under any pure stationary policies for the player, the state space does not contain more than one ergodic class.
  - ii) There exists a finite set $\mathcal{M} \subset E$ and a constant $\beta < 1$ such that:

$$\sum_{j \in \mathbf{I}} {}_{\mathcal{M}}p(i,a,b,j)\, \mu_j \leq \beta\, \mu_i \qquad \forall a, b, \forall i \in \mathbf{I} \tag{5}$$

    where ${}_{\mathcal{M}}p(i,a,b,j) = p(i,a,b,j)$ if j does not belong to the set $\mathcal{M}$, and is nul otherwise. (5) can be rewritten as $\|{}_{\mathcal{M}}P(u,v)\|_{\mu} \leq \beta$ for all pure stationary policies u and v.

Remark 3.1 *If assumptions $(\mathcal{B}1)$ and $(\mathcal{B}2)$ hold for some $\mathcal{M}$ and μ, then one can choose a state $0 \in \mathcal{M}$ and another μ' such that these assumptions hold with the set $\mathcal{M}' = \{0\}$ replacing $\mathcal{M}$, and μ' replacing μ, (see [23]). Therefore, we assume in the sequel, without loss of generality, that $\mathcal{M} = \{0\}$, for some state 0.*

Define

$$S(i,u,v) = \overline{\lim_{t \to \infty}} \frac{1}{t} E_i^{u,v} \sum_{s=0}^{t-1} r(I_s, A_s, B_s) \tag{6}$$

Theorem 3.1 *(see [7]) Suppose that assumptions $(\mathcal{B}1)$ and $(\mathcal{B}2)$ hold. Then*
(i) The stochastic game with expected average payoff criterion has a value,
(ii) There exists a unique solution pair (g, v), $g \in \mathbb{R}, \mathrm{v} \in \mathrm{R}^{\mathbf{I}}$ to the functional equation:

$$v(i) = \mathbf{Val}\left\{ r(i,a,b) - g + \sum_{j \in \mathbf{I}} p(i,a,b,j) v(j) \right\}_{a,b} \qquad i \in \mathbf{I}. \tag{7}$$

such that $|v|$ is finite and $v(0) = 0$.
(iii) g is the unique value of the stochastic game.
(iv) Let (u,v) be stationary policies such that $u(i), v(i)$ are optimal for the dummy game in the curly brackets in (7). Then, they are optimal for the stochastic game.
(v) $R(i) = \mathbf{Val}\{S(i,u,v)\}_{u,v} = g$.

4 State truncation and approximation

In the following approximating schemes, we modify the "limit" stochastic game in the following way. We Consider an increasing set of states $\mathbf{I}_1, \mathbf{I}_2, ...$ converging to $\mathbf{I}$, such that $0 \in \mathbf{I}_1$. The nth stochastic game is restricted to the set $\mathbf{I}_n$. In the game G_n, we modify the transition probabilities so as to eliminate all transitions outside the set $\mathbf{I}_n$. The two schemes will differ by the way that such transitions will be replaced. Introduce the following assumption

- $(\mathcal{B}3)$ $\quad \delta(r,n) = \sup_{\substack{i\in\mathbf{I}_r \\ a\in\mathbf{A}, b\in\mathbf{B}}} \sum_{j\notin\mathbf{I}_n} p(i,a,b,j) \to 0 \text{ as } n \to +\infty, \quad \forall r.$

Under the assumptions of our model (i.e. $(\mathcal{B}1)$-$(\mathcal{B}2)$), $(\mathcal{B}3)$ holds if $\mathbf{I}_n$ are finite sets $\forall n$, see [25].

4.1 Scheme I

In the game G_n, we modify the transition probabilities so as to eliminate all transitions outside the set $\mathbf{I}_n$; we replace transitions outside of $\mathbf{I}_n$ by transitions to state 0. Hence, $p^n(i,a,b,j)$ is defined by:

$$p^n(i,a,b,j) = \begin{cases} p(i,a,b,0) + \sum_{l\notin\mathbf{I}_n} p(i,a,b,l) & j = 0 \\ p(i,a,b,j) & j \neq 0, j \in \mathbf{I}_n \\ 0 & j \notin \mathbf{I}_n \end{cases} \tag{8}$$

$S_n(i,u,v,)$ is now defined as (6), where the expectation is taken with respect to the measure generated by the new transition probabilities (8). For $n \in \mathbb{N}$, let the pair $(g_n, v_n \in \bigvee_\mu)$ be the solutions of the dynamic programming equation :

$$\begin{aligned} &v_n(i) = \mathbf{Val}\left\{r(i,a,b) - g_n + \textstyle\sum_{j\in\mathbf{I}} p^n(i,a,b,j) v_n(j)\right\}_{a,b} \quad i \in \mathbf{I}_n, i \neq 0 \\ &v_n(0) = 0. \end{aligned} \tag{9}$$

for all n, g_n and v_n indeed exist and are unique since the assumptions $(\mathcal{B}1)$ and $(\mathcal{B}2)$ remain valid for this problem and theorem 3.1 applies. Moreover, we have that:

$$R_n(i) = \mathbf{Val}\{S_n(i,u,v)\}_{u,v} = g_n \qquad \forall i \in \mathbf{I}.$$

In order to prove the convergence of the state approximation scheme we introduce the following quantities:

- $\tau := \inf\{t \geq 1, I_t = 0\}$ is the time to reach state zero (with the convention that $\inf\{\emptyset\} = \infty$).
- $w(i,u,v)$:= The total cost to reach zero from state i when policies u and v are used

$$w(i,u,v) = E_i^{u,v} \sum_{s=0}^{\tau} r(I_s, A_s, B_s)$$

  which can be rewritten in a vector form as:

$$w(u,v) = E_i^{u,v} \sum_{s=0}^{\infty} [{}_0p(u,v)]^s r(u,v).$$

- $w^n(i,u,v)$:= The total cost to reach zero from state i when policies u and v are used, when the transition probabilities are replaced by (8).
- $\overline{\tau}(i,u,v), \overline{\tau}^n(i,u,v)$:= The expectations of τ when using the original transition probabilities, and when using those in (8), respectively.

We note that $w(\cdot,u,v)$, $w^n(\cdot,u,v)$, $\overline{\tau}(\cdot,u,v)$ and $\overline{\tau}^n(\cdot,u,v)$ are uniformly μ bounded. Indeed,

$$\|w(u,v)\|_\mu \leq \sum_{s=0}^{\infty} \left[\|{}_0p(u,v)\|_\mu\right]^s \|r(u,v)\|_\mu \leq \frac{M}{1-\beta}$$

with the same bound for $w^n(u,v)$. Similarly, $\|\overline{\tau}(u,v)\|_\mu$ and $\|\overline{\tau}^n(u,v)\|_\mu$ are bounded by $(1-\beta)^{-1}$. It is easily seen that $w(\cdot,u,v)$, $w^n(\cdot,u,v)$, $\overline{\tau}(\cdot,u,v)$ and $\overline{\tau}^n(\cdot,u,v)$ are the unique solution in $\bigvee_\mu$ of the fixed point equations

$$w(i,u,v) = r(i,u,v) + \sum_{j\neq 0} p(i,u,v,j) w(j,u,v) \tag{10}$$

$$w^n(i,u,v) = \begin{cases} r(i,u,v) + \sum_{j\neq 0} p^n(i,u,v,j) w^n(j,u,v), \text{ for } i \in I_n \\ w(i,u,v) \text{ for } i \notin I_n \end{cases} \tag{11}$$

$$\overline{\tau}(i,u,v) = 1 + \sum_{j\neq 0} p(i,u,v,j) \overline{\tau}(j,u,v) \tag{12}$$

$$\overline{\tau}^n(i,u,v) = \begin{cases} 1 + \sum_{j\neq 0} p^n(i,u,v,j) \overline{\tau}^n(j,u,v), \text{ for } i \in I_n \\ \overline{\tau}(i,u,v) \text{ for } i \notin I_n \end{cases} \tag{13}$$

The uniqueness follows from the fact that the above equations are contracting due to $(\mathcal{B}2)$. Note that functions $w(.,u,v)$ and $w^n(.,u,v)$ are μ-bounded for all pair (u,v) on every subset J of $\mathbf{I}$. Since both $w(0,u,v)$ and $\overline{\tau}(0,u,v)$ are both non-zero and finite, it follows (see Chung [12] p. 91-92) that the expected average cost is given by the following ratio between the the total cost and the expected hitting time of state zero

$$S(i,u,v) = \frac{w(0,u,v)}{E\tau(0,u,v)} \quad \text{and } S_n(i,u,v) = \quad \frac{w^n(0,u,v)}{E\tau^n(0,u,v)}. \tag{14}$$

Theorem 4.1 *Assume $(\mathcal{B}1)$-$(\mathcal{B}3)$. All statements of Theorem* 2.1 *hold, where S_n and S are the expected average payoffs defined in* (6)*, with the transition probabilities p and p^n (defined in (8)) respectively.*

Proof. Fix some initial state i. We use Theorem 2.1. We begin by establishing conditions $(\mathcal{A}1)$ and $(\mathcal{A}2)$. Since $U = U_n$ and $V_n = V$ for each n, it suffices to show that $S_n(u,v) := S_n(i,u,v)$ converges to $S(u,v) := S(i,u,v)$ uniformly on $\mathbf{I}$. Hence, we set π_n^1, π_n^2, σ_n^1 and σ_n^2 to be identity.
Let J be a given subset of $\mathbf{I}$, and (u,v) a pair of strategies. To avoid cumbersome notations we will write $w(.)$ (resp $w^n(.)$) instead of $w(.,u,v)$ (resp $w^n(.,u,v)$). We first want to prove that

$$\lim_{n\to+\infty} \parallel w^n - w \parallel_\mu^J = 0$$

once we show that, one obtains in the same way that $\lim_{n\to+\infty} \parallel E\tau - E\tau^n \parallel_\mu^J = 0$, and the uniform convergence of $S_n(u,v)$ to $S(u,v)$ now follows from (14).

We use an idea introduced by Cavazos-Cadena [11] and used in [25] for a similar problem. Fix ϵ arbitrarily small, and define the sequence g_k in the following way. $g_0 = \min\{m : J \subset \mathbf{I}_m\}$ and recursively,

$$g_k = g(\epsilon, g_{k-1}), \quad g(\epsilon,r) = \min\{m : \delta(r,m) \leq \epsilon\},$$

where δ is defined in $(\mathcal{B}3)$. Due to assumption $(\mathcal{B}3)$ this sequence is well defined, and for all k, g_k is finite. Let ν a given integer, define also

$$m^\nu(\epsilon) = \max\{g_m, m = 0,1...,\nu\}$$

Let $n \geq m^\nu(\epsilon)$, $i \in J$.
Let us now compute $\parallel w^n - w \parallel_\mu^J$. We obviously have that $\parallel w^n - w \parallel_\mu^J \leq \parallel w^n - w \parallel_\mu^{I_0}$

since $J \subset I_0$, and for $i \in \mathbf{I}_0$,

$$\begin{aligned}
\frac{1}{\mu_i}|w^n(i) - w(i)| &= \frac{1}{\mu_i}|\sum_{j\neq 0} p^n(i,u,v,j)w^n(j) - p(i,u,v,j)w(j)| \\
&\leq \frac{1}{\mu_i}\sum_{j\in \mathbf{I}_{g_1}\setminus\{0\}} |p^n(i,u,v,j)w^n(j) - p(i,u,v,j)w(j)| \\
&\quad + \frac{1}{\mu_i}\sum_{j\in \mathbf{I}_n\setminus\mathbf{I}_{g_1}} |p^n(i,u,v,j)w^n(j) - p(i,u,v,j)w(j)| \\
&\leq \sum_{j\in \mathbf{I}_{g_1}\setminus\{0\}} \frac{p(i,u,v,j)\mu_j}{\mu_i}\frac{|w^n(j) - w(j)|}{\mu_j} \\
&\quad + \frac{1}{\mu_i}\sum_{j\in \mathbf{I}_n\setminus\mathbf{I}_{g_1}} p(i,u,v,j)|w^n(j) - w(j)|
\end{aligned}$$

In the last inequality the first term can be bounded by $\beta \parallel w - w^n \parallel_\mu^{\mathbf{I}_{g_1}}$ because of assumption ($\mathcal{B}2$,ii), and the second by $2\frac{M}{1-\beta}\epsilon$, because of the definition of the sequence I_{g_k}, since i belongs to I_{g_0} plus ($\mathcal{B}3$), and since w^n and w are bounded by $\frac{M}{1-\beta}$. We obtain

$$\parallel w^n - w \parallel_\mu^{\mathbf{I}_0} \leq \beta \parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_1}} + 2\frac{M}{1-\beta}\epsilon$$

Exactly in the same way we get for $k \leq m^\nu(\epsilon)$

$$\parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_k}} \leq \beta \parallel w^n - w \parallel_\mu^{\mathbf{I}_{g_{k+1}}} + 2\frac{M}{1-\beta}\epsilon$$

and finally

$$\parallel w^n - w \parallel_\mu^{J} \leq \beta^\nu \frac{2M}{1-\beta} + \frac{2M\epsilon}{1-\beta}\left(\frac{1-\beta^\nu}{1-\beta}\right). \tag{15}$$

Since ν can be chosen arbitrarily large when n tends to infinity, and β is strictly lower than 1, this bound can be as small as needed for n large enough. This establishes ($\mathcal{A}1$)-($\mathcal{A}2$) in Theorem 2.1. It follows from [7] that S is a continuous function of u and v, which implies ($\mathcal{A}3$)-($\mathcal{A}4$) of Theorem 2.1. This establish the proof. ∎

4.2 Scheme II

In the previous scheme, we replaced transitions outside of $\mathbf{I}_n$ by transitions to state 0. In some applications this may be undesirable; this is the case when the games with truncated space describe real problems that we wish to approximate by some game with an infinite state space. To illustrate this, consider a queue with a finite length L, and assume that the state is the number of customers in the queue. Then, typically, if a transition from state L to state $L+1$ were possible in the case of infinite queue, then in the problem with truncated state space, which corresponds to a finite queue, it is replaced by a transition from L to L. In the previous scheme, it would be replaced by a transition to state 0. This would be especially undesirable, since in queueing problems, we usually have the property of transitions to closest neighbors: from each state, only finitely many neighboring states can be reached in one step. So, having a transition from state L to 0 does not describe a realistic model of a finite queue.

In the following scheme, we consider a truncation that is adapted to problems with the property of closest neighbor transitions, and that replace transitions outside of $\mathbf{I}_n$ by transitions to the "boundaries". We thus assume

($\mathcal{B}4$): For any i, a, b, $p(i,a,b,\bullet)$ has finite support (that may depend on i, a, b),

and define the following approximation scheme.

(i) For all $m = 1, 2, ...$, the state space is decomposed in two disjoint classes of states: E^m, which contains a finite number of states, and T^m.

(ii) For all m large enough the following holds: under any stationary policy u, E^m contains one recurrent class (plus possibly some transient states), T^m is a transient class, and absorption into the positive recurrent class takes place in finite expected time from any initial state.

(iii) $E_m \subset E_{m+1}$, $m = 1, ...$; $E_\infty = \lim_{m\to\infty} E_m = \mathbf{I}$.

(iv) There is some partial order on $\mathbf{I}$, and we assume that the following holds:

$$p^n(i,a,b,j) \begin{cases} = p(i,a,b,j) & i \in E_m,\ j \in E_{m-1} \\ \geq p(i,a,b,j) & i \in E_m,\ j \in E_m \backslash E_{m-1} \\ = 0 & i \in E_m,\ j \notin E_m \\ = 1\{j = 0\} & i \notin E_m \end{cases} \tag{16}$$

where 0 is some arbitrary state in E_0. Moreover, for every $m > 1$ and each $j \in E_m \backslash E_{m-1}$ and $i \in E_{m-1}$, we have $i \leq j$ w.r.t. the partial order on $\mathbf{I}$.

Again, $S_n(i,u,v,)$ is defined as (6), where the expectation is taken with respect to the measure generated by the new transition probabilities (16).

Theorem 4.2 *Assume ($\mathcal{B}1$), ($\mathcal{B}2$) and ($\mathcal{B}4$), and consider the above finite approximation scheme. All statements of Theorem 2.1 hold, where S_n and S are the expected average payoffs defined in* (6)*, with the transition probabilities p and p^n (defined in (16)) respectively.*

Proof. It follows from the proof of Theorem 5.1 in [2] (see eq. (5.2)) that $S_n(u,v)$ converge to $S(u,v)$ uniformly in all stationary policies. This implies assumptions ($\mathcal{A}1$) and ($\mathcal{A}2$). Assumptions ($\mathcal{A}3$) and ($\mathcal{A}4$) relate only to the limit game, and therefore the proof is the same as in the previous Section. The theorem now follows from Theorem 2.1. ∎

5 Convergence of the discounted cost to the average cost

Conditions for the convergence of the value and equilibrium policies for discounted cost stochastic games to those of the average cost game are well known, see e.g. [14]. These were extended recently to unbounded cost (see [7, 21]). Theorem 2.1 enables us not only to obtain an alternative proof for the above convergence of the values and policies, but also to obtain new robustness results, see Theorem 5.2 below.

Define the **β-discounted game payoff**

$$S_\beta(i,u,v) = (1-\beta)E_i^{u,v}\sum_{t=0}^{\infty}\beta^t r(I_t,A_t,B_t) \tag{17}$$

The following was proved in [7] Theorem 3.4:

Theorem 5.1 *Assume ($\mathcal{B}1$) and ($\mathcal{B}2$). Then*
(i) A value $R_\beta(i)$ exists for the discounted cost.
(ii) Optimal stationary policies exist for both players for any discount factor $0 < \beta < 1$ (they are said to be β-optimal).
(iii) Any limit-point (as β tends to one) of β-optimal stationary policies is expected average optimal; moreover, the value of the discounted games converge to the value of the expected average game.

Theorem 5.2 *(i) Let (u^*,v^*) be any stationary policy pair which is expected average optimal. Then for any $\epsilon > 0$, (u^*,v^*) is ϵ optimal for the β-discounted cost, for all β sufficiently close to 1, and for all $u \in U$, $v \in V$,*

$$\overline{\lim_{\beta\to 1}}[S_\beta(i,u,v^*) - S_\beta(i,u^*,v^*)] \le 0, \qquad \underline{\lim_{\beta\to 1}}[S_\beta(i,u^*,v) - S_\beta(i,u^*,v^*)] \ge 0.$$

(ii) For any $\epsilon > 0$ there exsists some $\beta_0 < 1$ such that for any $\beta_0 \leq \beta < 1$ and any stationary pair u^β, v^β which are β-optimal, (u^β, v^β) is ϵ-optimal for the expected average game.

Proof. It is sufficient to prove that $(\mathcal{A}1)$ and $(\mathcal{A}2)$ in Theorem 2.1 hold. This follows indeed from the fact that

$$\lim_{\beta \to 1} \|S_\beta(\cdot, u, v) - S(\cdot, u, v)\|_\mu = 0$$

uniformly over all stationary policies u and v, see [2] p. 166. ∎

6 Stochastic games with complete information

In Sections 4, 5 and in [25] we described several approximation problems in stochastic games, where we had a value in the limit game. In all those cases, we considered (without loss of optimality) the (randomized) stationary or the (randomized) Markov policies. If we now restrict to pure stationary or pure Markov policies, the corresponding games will not have in general a value.

This restriction to pure strategies is equivalent to playing stochastic games with complete information, see [15, 18], in which the information structure is slightly different than the one we considered in Section 3. The information available for both players at time t is the same as in the standard model described in Section 3; the second player, however, has in addition to that, the information on the action chosen at time t by the first player. The equivalence between a stochastic game with complete information and a standard stochastic game with restriction to pure policies is the following. First, if a stochastic game with complete information is played, then by standard arguments, the players may restrict to pure strategies, without loss of optimality. Suppose the state is x at time t. In any game, the second player, chooses a policy as a function of the policy of the first player. Since the first player restricts to pure strategies, knowing that the state is x and knowing the strategy of player 1, enables player 2 to know what action will be played at time t by player 1. (Note that if player 1 did not restrict to pure strategies, then this argument would not hold). Hence, the standard stochastic game has the same information structure as the one with complete information.

Since we established conditions $(\mathcal{A}1)$ and $(\mathcal{A}2)$ for all the problems considered in Sections 4, 5 and in [25], they hold in particular if we restrict to pure strategies (or equivalently, if games with complete information are player instead). Therefore,

the convergence of policies and values in Theorems 2.2 and 2.3 hold for all these problems.

Applying Theorem 2.4 is more delicate, since only in special cases, can we define a topology over the space of responses of player 2 such that assumptions ($\mathcal{A}3$) and ($\mathcal{A}4$) holds (as opposed to standard stochastic games, where ($\mathcal{A}3$) and ($\mathcal{A}4$) need to hold for policies, and not for responses). In case that the action space available to player 1 is finite, one may identify the class of pure stationary response strategies of player 2 (corresponding to pure stationary policies of player 1) with the set of functions $\mathbf{I} \rightarrow \mathbf{A} \rightarrow \mathbf{B}$, endowed with the weak topology. The continuity assumptions ($\mathcal{A}3$) and ($\mathcal{A}4$) can now be established using arguments as in Remark 3.1 in [25] and [7].

7 A finite approximation of strategy sets

Another type of approximation that arises in dynamic games is the countable or finite approximation of strategy sets. This step is necessary when we want to perform numerical computations, and when the strategy sets are infinite or continuous, or both.

Let $\mathcal{U}$ and $\mathcal{V}$ be metric sets of policies for players one and two, and let $S(u,v)$ correspond to the cost associated to the pair of strategies $u \in \mathcal{U}, v \in \mathcal{V}$. Introduce the following sets of strategies: $U \subset \mathcal{U}$ and $V \subset \mathcal{V}$, and the sequences $\{U_n\}_{n\in\mathbb{N}} \subset \mathcal{U}$ and $\{V_n\}_{n\in\mathbb{N}} \subset \mathcal{V}$. U_n and V_n are assumed to be countable or finite sets of policies.

Theorem 7.1 *Suppose that*
($\mathcal{A}$'1): $\lim_{n\rightarrow+\infty} U_n = U$ *and* $\lim_{n\rightarrow+\infty} V_n = V$, *in the Hausdorf topology sense,*
($\mathcal{A}$'3): $S(u,v)$ *is a lower semicontinuous function in* $u \in U$ *uniformly in* $v \in V$,
($\mathcal{A}$'4): $S(u,v)$ *is a upper semicontinuous function in* $v \in V$ *uniformly in* $u \in U$,

Then the conclusion of theorem 2.1 hold (where $\overline{R}_n$ *and* $\underline{R}_n$ *are defined in (1), and* R *is defined with respect to the policy sets* U *and* V, *and for* $\epsilon_1 = 0$*).*

Proof. We need to prove that under the set of assumptions ($\mathcal{A}$'1), ($\mathcal{A}$'3) and ($\mathcal{A}$'4) the hypothesis of theorem 2.1 are satisfied. ($\mathcal{A}$'3) and ($\mathcal{A}$'4) directly imply ($\mathcal{A}3$) and ($\mathcal{A}4$). We shall only prove that ($\mathcal{A}1$) holds as the proof of ($\mathcal{A}2$) is identical. Choose

any ϵ_0 and introduce some sequence of functions:

$$\pi_n^1 : U_n \longrightarrow U, \text{ such that } d(u_n, \pi_n^1(u_n)) < \inf_{u \in U} d(u_n, u) + \epsilon_0$$

and

$$\sigma_n^2 : V \longrightarrow V_n, \text{ such that } d(\sigma_n^2(v), v) < \inf_{v_n \in V_n} d(v_n, v) + \epsilon_0.$$

By $(\mathcal{A}'\ 1)$, for all ϵ_2 there exists $N_1 = N_1(\epsilon_2)$ such that for all $n > N_1$,

$$\max \left(\sup_{u_n \in U_n} \inf_{u \in U} d(u_n, u) \ ; \ \sup_{u \in U} \inf_{u_n \in U_n} d(u_n, u) \right) \leq \epsilon_2,$$

that is for all $u_n \in U_n$,

$$\inf_{u \in U} d(u_n, u) \leq \epsilon_2,$$

so for all $u_n \in U_n$,

$$d(u_n, \pi_n^1(u_n)) \leq \epsilon_0 + \epsilon_2. \tag{18}$$

Similarly, for all ϵ_3 there exists $N_3 = N_3(\epsilon_3)$ such that for all $n > N_3$ and for all $v \in V$,

$$d(v, \sigma_n^2(v)) \leq \epsilon_0 + \epsilon_3. \tag{19}$$

To prove $(\mathcal{A}1)$ we shall show that for all ϵ, there exists $N = N(\epsilon)$ such that for all $n > N$, and for all $u_n \in U_n$, $v \in V$ we have :

$$S(u_n, \sigma_n^2(v)) - S(\pi_n^1(u_n), v) = S(u_n, \sigma_n^2(v)) - S(u_n, v) + S(u_n, v) - S(\pi_n^1(u_n), v) < \epsilon. \tag{20}$$

Since S is upper semicontinuous in v uniformly in u, there exists η_1 such that if $d(\sigma_n^2(v), v) \leq \eta_1$, $S(u_n, \sigma_n^2(v)) - S(u_n, v) \leq \epsilon/2$. Choose ϵ_0 and ϵ_3 such that $\epsilon_0 + \epsilon_3 = \eta_1$ in (19), there exists N_1 such that for all $n > N_1$,

$$S(u_n, \sigma_n^2(v)) - S(u_n, v) \leq \frac{\epsilon}{2}. \tag{21}$$

Similarly, it follows from (18) and since S is lower semicontinuous in u uniformly in v that there exists N_2 such that for all $n > N_2$

$$S(u_n, v) - S(\pi_n^1(u_n), v) \leq \frac{\epsilon}{2}. \tag{22}$$

(21) and (22) imply (20) by choosing $N = \sup(N_1, N_2)$, which concludes the proof. ■

Remark 7.1 *In many applications (e.g. [9, 22]), the strategy sets are compact. Hence it suffices to require in ($\mathcal{A}$' 3) and ($\mathcal{A}$' 4) the semi-continuity properties; the uniform semi-continuity is then a consequence of the compactness of the strategy sets.*

As an application of the Theorem 7.1, we present the following continuous time differential pursuit evasion game by Bernhard and Shinar [9, 22]. We shall use the same notation as in [9, 22]. The game is governed by a differential equation

$$\frac{dx}{dt} = f(x,a,b),\ x \in \mathbf{I}, a \in \mathbf{A}, b \in \mathbf{B}$$

where $\mathbf{A}$ and $\mathbf{B}$ are bounded subsets of $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$ respectively, $\mathbf{I}$ is a domain of $\mathbb{R}^n$. Some regularity and growth conditions on f insure the existence of the solution of the differential equation over $(0,\infty)$ for every pair of measurable functions $a(\cdot)$ (and $b(\cdot)$) from $(0,\infty)$ to $\mathbf{A}$ (respectively, $\mathbf{B}$). The players have access to noisy partial information

$$y_a = h_a(x,w), \qquad y_b = h_b(x,w)$$

where w is a noise and h_a, h_b are globally Lipschitz over $\mathbf{I}$. They are restricted to using feedback strategies $(a(t) = \delta_1(y_a(t)), (b(t) = \delta_2(y_b(t)))$ Lipschitz continuous, then the set of strategies is compact in the topology of the uniform convergence. It is assumed that the noise model and the solution concept of the differential equations are such that the payoff P (the expected value of a continuous function of closest approach) is a continuous function of the strategies for the topology of uniform convergence.

If Ω_1 and Ω_2 are compact metric strategies spaces and Δ_1 and Δ_2 are closed subsets of Ω_1 and Ω_2 respectively and $U = \Pi(\Delta_1)$ and $V = \Pi(\delta_2)$ are the sets of probability measures over Δ_1 and Δ_2, we know that there exist optimal mixed strategies that achieve the value

$$V(\Delta_1,\Delta_2) = \min_{u\in U}\max_{v\in V} J(u,v) = \int_{\Delta_1}\int_{\Delta_2} P(\delta_1,\delta_2)du(\delta_1)dv(\delta_2)$$

Bernhard and Shinar establish the convergence of the values of some approximating problems to the value of the original one.

We show that, in fact, all the statments concerning the convergence of policies in Theorem 2.1 also hold (with $\epsilon_1 = 0$). In [9], the continuity of $V(.,.)$ is proved, i.e. ($\mathcal{A}$' 3) and ($\mathcal{A}$' 4) are established. They present a finite approximation of this problem by considering finite subsets of Δ_i $i = 1,2$ that converge to Δ_i in the Hausdorf topology (and thus, ($\mathcal{A}$' 1) holds).

References

[1] E. Altman, "Denumerable Constrained Markov Decision Problems and Finite Approximations", *Math. of Operations Research*, **19**, No. 1, pp. 169-191, 1994.

[2] E. Altman, "Asymptotic Properties of Constrained Markov Decision Processes", *Zeitschrift für Operations Research*, Vol. 37, Issue 2, pp. 151-170, 1993.

[3] E. Altman, "Flow control using the theory of zero-sum Markov games," *IEEE-AC*, April, 1994.

[4] E. Altman, "A Markov game approach for optimal routing into a queueing network", INRIA report No. 2178, submitted, 1994.

[5] E. Altman, "Non zero-sum stochastic games in admission, service and routing control in queueing systems", submitted to *Journal of Applied Probability.*

[6] E. Altman and A. Hordijk, "Zero-sum Markov games and worst-case optimal control of queueing systems", invited paper, submitted to *QUESTA*, special issue on optimization of queueing systems, 1994. Research report No. TW-94-01, University of Leiden.

[7] E. Altman, A. Hordijk and F. M. Spieksma, "Contraction conditions for average and α-discounted optimality in countable state Markov games with unbounded rewards", submitted to *MOR*, 1994.

[8] E. Altman and G. Koole, "Stochastic Scheduling Games with Markov Decision Arrival Processes", *Journal Computers and Mathematics with Appl.*, third special issue of "Differential Games", pp. 141-148, 1993.

[9] P. Bernhard and J. Shinar, " On finite Approximation of a Game Solution with Mixed Strategies", *Appl. Math. Lett.* **3** No. 1, pp 1-4, 1990.

[10] V. Borkar and M. K. Ghosh, "Denumerable state stochastic games with limiting average payoff", *To appear in JOTA*, 1992.

[11] R. Cavazos-Cadena, "Finite-state approximations for denumerable state discounted Markov Decision Processes", *J. Applied Mathematics and Optimization* **14** pp. 27-47, 1986.

[12] K. L. Chung, *Markov chains with stationary transition probabilities*, 2nd edition, Springer Verlag, New York, 1967.

[13] R. Dekker and A. Hordijk, "Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards", *Mathematics of Operations Research*, **13**, pp. 395-421, 1988.

[14] A. Federgruen, "On N-Person Stochastic Games with denumerable state space", *Adv. Appl. Prob.* **10**, pp. 452-471, 1978.

[15] D. Gillette, "Stochastic games with zero stop probabilities", M. Dresher, A. W. Tucker, P. Wolfe, eds., Princeton University Press, Princeton, 1957, pp. 179-187.

[16] O. Hernandez-Lerma, "Finite state approximations for denumerable multidimensional - state discounted Markov decision processes", *J. Mathematical Analysis and Applications* **113** pp. 382-389, 1986.

[17] O. Hernandez-Lerma, *Adaptive Control of Markov Processes*, Springer Verlag, 1989.

[18] H. W. Küenle, *Stochastiche Spiele und Entscheidungesmodelle*, Tebuner-Texte, Band 89, 1986.

[19] A. S. Nowak, "On zero-sum stochastic games with general state space, I", *Probability and Mathematical Statistics* IV, No. 1, pp. 13-32, 1984.

[20] A. S. Nowak, "Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space", *JOTA* **45**, No. 4, pp. 592-602, 1985.

[21] L. I. Sennott, "Zero-sum stochastic games with unbounded costs: discounted and average cost cases" *Technical report*, 1992, to appear in *ZOR*.

[22] J. Shinar and I. Forte, "On the optimal pure strategy sets for a missile guidance law synthesis", *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, 1986.

[23] F. M. Spieksma, *Geometrically Ergodic Markov Chains and the Optimal Control of Queues*, Ph.D. thesis, 1990, Leiden University (available on request from the author).

[24] L. C. Thomas and D. Stengos, "Finite State Approximation Algorithms for Average Cost Denumerable State Markov Decision Processes", *OR Spectrum*, **7**, pp. 27-37, 1985.

[25] M. Tidball and E. Altman, "Approximations in dynamic zero-sum games, I", INRIA report No. 2166, Submitted to *SIAM J. Control and Optimization.*

[26] D. J. White, "Finite State Approximations for Denumerable State Infinite Horizon Discounted Markov Decision Processes", *J. Mathematical Analysis and Applications* **74**, pp. 292-295, 1980.

[27] D. J. White, "Finite State Approximations for Denumerable State Infinite Horizon Discounted Markov Decision Processes with Unbounded Rewards", *J. Mathematical Analysis and Applications* **86**, pp. 292-306, 1982.

[28] W. Whitt, "Approximations of Dynamic Programs, I", *Mathematics of Operations Research*, Vol. 3 No. 3, pp. 231-243, 1978.

[29] W. Whitt, "Representation and Approximation of Noncooperative Sequential Games", *SIAM J. Control and Opt.*, Vol 18 No 1, pp. 33-43, 1980.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399